from unwanted reaction, the protecting groups block a reactive end of the monomer to prevent self-polymerization. For instance, attachment of a protecting group to the amino terminus of an activated amino acid, such as an N-hydroxysuccinimide-activated ester of the amino acid, prevents the amino terminus of one monomer from reacting with the activated ester portion of another during peptide synthesis. Alternatively, the protecting group may be attached to the carboxyl group of an amino acid to prevent reaction at this site. Most protecting groups can be attached to either the amino or the carboxyl group of an amino acid, and the nature of the chemical synthesis will dictate which reactive group will require a protecting group. Analogously, attachment of a protecting group to the 5'-hydroxyl group of a nucleoside during synthesis using for example, phosphate-triester coupling chemistry, prevents the 5'-hydroxyl of one nucleoside from reacting with the 3'-activated phosphate-triester of another.

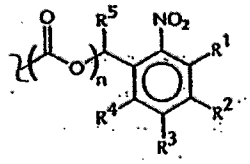
Regardless of the specific use, protecting groups are employed to protect a moiety on a molecule from reacting with another reagent. Protecting groups of the present invention have the following characteristics: they prevent selected reagents from modifying the group to which they are attached; they are stable (that is, they remain attached to the molecule) to the synthesis reaction conditions; they are removable under conditions that do not adversely affect the remaining structure; and once removed, do not react appreciably with the surface or surface-bound oligomer. The selection of a suitable protecting group will depend, of course, on the chemical nature of the monomer unit and oligomer, as well as the specific reagents they are to protect against.

In a preferred embodiment, the protecting groups are photoactivatable. The properties and uses of photoreactive protecting compounds have been reviewed. See, McCray et al., Ann. Rev. of Biophys. and Biophys.

Chem. (1989) 18:239-270, which is incorporated herein by reference. Preferably, the photosensitive protecting groups will be removable by radiation in the ultraviolet (UV) or visible portion of the electromagnetic spectrum. More preferably, the protecting groups will be removable by radiation in the near UV or visible portion of the spectrum. In some embodiments, however, activation may be performed by other methods such as localized heating, electron beam lithography, laser pumping, oxidation or reduction with microelectrodes, and the like. Sulfonyl compounds are suitable reactive groups for electron beam lithography. Oxidative or reductive removal is accomplished by exposure of the protecting group to an electric current source, preferably using microelectrodes directed to the predefined regions of the surface which are desired for activation. Other methods may be used in light of this disclosure.

Many, although not all, of the photoremovable protecting groups will be aromatic compounds that absorb near-UV and visible radiation. Suitable photoremovable protecting groups are described in, for example, McCray et al., Patchornik, J. Amer. Chem. Soc. (1970) 92:6333, and Amit et al., J. Org. Chem. (1974) 39:192, which are incorporated herein by reference.

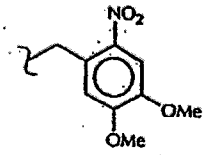
A preferred class of photoremovable protecting groups has the general formula:



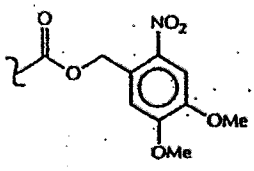
where R^1, R^2, R^3, and R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents

(i.e., R^1-R^2, R^2-R^3, R^3-R^4) are substituted oxygen groups that together form a cyclic acetal or ketal; R^5 is a hydrogen atom, a alkoxyl, alkyl, hydrogen, halo, aryl, or alkenyl group, and n = 0 or 1.

A preferred protecting group, 6-nitroveratryl (NV), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 are each a methoxy group, R^1, R^4 and R^5 are each a hydrogen atom, and n = 0:

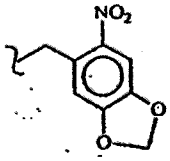


A preferred protecting group, 6-nitroveratryloxycarbonyl (NVOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 are each a methoxy group, R^1, R^4 and R^5 are each a hydrogen atom, and n = 1:

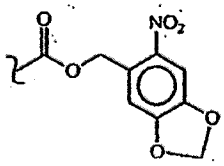


Another preferred protecting group, 6-nitropiperonyl (NP), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1, R^4 and R^5 are each a hydrogen atom, and n = 0:

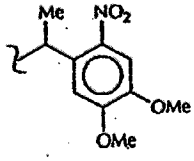
40



Another preferred protecting group, 6-nitropiperonyloxycarbonyl (NPOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1, R^4 and R^5 are each a hydrogen atom, and n = 1:

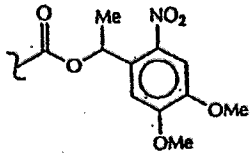


A most preferred protecting group, methyl-6-nitroveratryl (MeNV), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 are each a methoxy group, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and n = 0:

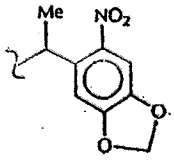


Another most preferred protecting group, methyl-6-nitroveratryloxycarbonyl (MeNVOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 are each a methoxy

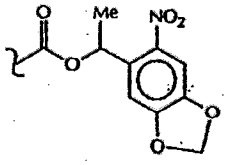
group, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and n = 1:



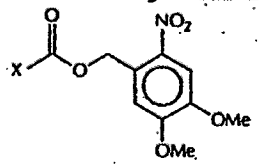
Another most preferred protecting group, methyl-6-nitropiperonyl (MeNP), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and n = 0:



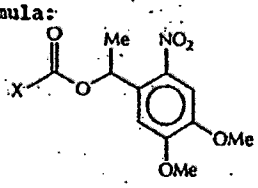
Another most preferred protecting group, methyl-6-nitropiperonyloxycarbonyl (MeNPOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and n = 1:



A protected amino acid having a photoactivatable oxycarbonyl protecting group, such NVOC or NPOC or their corresponding methyl derivatives, MeNVOC or MeNPOC, respectively, on the amino terminus is formed by acylating the amine of the amino acid with an activated oxycarbonyl ester of the protecting group. Examples of activated oxycarbonyl esters of NVOC and MeNVOC have the general formula:



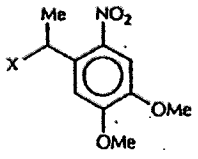
NVOC-X



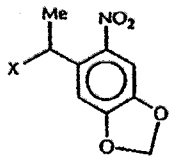
MeNVOC-X

where X is halogen, mixed anhydride, phenoxy, p-nitrophenoxy, N-hydroxysuccinimide, and the like.

A protected amino acid or nucleotide having a photoactivatable protecting group, such as NV or NP or their corresponding methyl derivatives, MeNV or MeNP, respectively, on the carboxy terminus of the amino acid or 5'-hydroxy terminus of the nucleotide, is formed by acylating the carboxy terminus or 5'-OH with an activated benzyl derivative of the protecting group. Examples of activated benzyl derivatives of MeNV and MeNP have the general formula:



MeNV-X



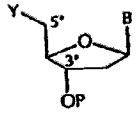
MeNP-X

where X is halogen, hydroxyl, tosyl, mesyl, trifluormethyl, diazo, azido, and the like.

Another method for generating protected monomers is to react the benzylic alcohol derivative of the protecting group with an activated ester of the

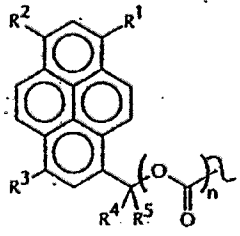
monomer. For example, to protect the carboxyl terminus of an amino acid, an activated ester of the amino acid is reacted with the alcohol derivative of the protecting group, such as 6-nitroveratrol (NVOH). Examples of activated esters suitable for such uses include halo-formate, mixed anhydride, imidazoyl formate, acyl halide, and also includes formation of the activated ester in situ the use of common reagents such as DCC and the like. See Atherton et al. for other examples of activated esters.

A further method for generating protected monomers is to react the benzylic alcohol derivative of the protecting group with an activated carbon of the monomer. For example, to protect the 5'-hydroxyl group of a nucleic acid, a derivative having a 5'-activated carbon is reacted with the alcohol derivative of the protecting group, such as methyl-6-nitropiperonol (MePyROH). Examples of nucleotides having activating groups attached to the 5'-hydroxyl group have the general formula:



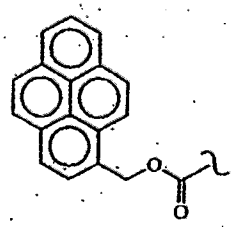
where Y is a halogen atom, a tosyl, mesyl, trifluoromethyl, azido, or diazo group, and the like.

Another class of preferred photochemical protecting groups has the formula:

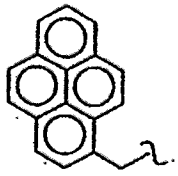


where R^1, R^2, and R^3 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfanates, sulfido or phosphido group, R^4 and R^5 independently are a hydrogen atom, an alkoxy, alkyl, halo, aryl, hydrogen, or alkenyl group, and n = 0 or 1.

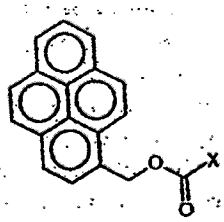
A preferred protecting group, 1-pyrenylmethyloxycarbonyl (PyROC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^1 through R^5 are each a hydrogen atom and n = 1:



Another preferred protecting group, 1-pyrenylmethyl (PyR), which is used for protecting the carboxy terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^1 through R^5 are each a hydrogen atom and n = 0:

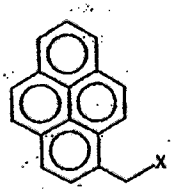


An amino acid having a pyrenylmethyloxycarbonyl protecting group on its amino terminus is formed by acylation of the free amine of amino acid with an activated oxycarbonyl ester of the pyrenyl protecting group. Examples of activated oxycarbonyl esters of PyROC have the general formula:



where X is halogen, or mixed anhydride, p-nitrophenoxy, or N-hydroxysuccinimide group, and the like.

A protected amino acid or nucleotide having a photoactivatable protecting group, such as PyR, on the carboxy terminus of the amino acid or 5'-hydroxy terminus of the nucleic acid, respectively, is formed by acylating the carboxy terminus or 5'-OH with an activated pyrenylmethyl derivative of the protecting group. Examples of activated pyrenylmethyl derivatives of PyR have the general formula:



where X is a halogen atom, a hydroxyl, diazo, or azido group, and the like.

Another method of generating protected monomers is to react the pyrenylmethyl alcohol moiety of the protecting group with an activated ester of the monomer. For example, an activated ester of an amino acid can be reacted with the alcohol derivative of the protecting group, such as pyrenylmethyl alcohol (PyROH), to form the protected derivative of the carboxy terminus of the amino acid. Examples of activated esters include halo-formate, mixed anhydride, imidazoyl formate, acyl halide, and also

includes formation of the activated ester in situ and the use of common reagents such as DCC and the like.

Clearly, many photosensitive protecting groups are suitable for use in the present invention.

In preferred embodiments, the substrate is irradiated to remove the photoremovable protecting groups and create regions having free reactive moieties and side products resulting from the protecting group. The removal rate of the protecting groups depends on the wavelength and intensity of the incident radiation, as well as the physical and chemical properties of the protecting group itself. Preferred protecting groups are removed at a faster rate and with a lower intensity of radiation. For example, at a given set of conditions, MeNVOC and MeNPOC are photolytically removed from the N-terminus of a peptide chain faster than their unsubstituted parent compounds, NVOC and NPOC, respectively.

Removal of the protecting group is accomplished by irradiation to liberate the reactive group and degradation products derived from the protecting group. Not wishing to be bound by theory, it is believed that irradiation of an NVOC- and MeNVOC-protected oligomers occurs by the following reaction schemes:

NVOC-AA -> 3,4-dimethoxy-6-nitrosobenzaldehyde + CO_2 + AA
MeNVOC-AA-> 3,4-dimethoxy-6-nitrosoacetophenone + CO_2 + AA

where AA represents the N-terminus of the amino acid oligomer.

Along with the unprotected amino acid, other products are liberated into solution: carbon dioxide and a 2,3-dimethoxy-6-nitrosophenylcarbonyl compound, which can react with nucleophilic portions of the oligomer to form unwanted secondary reactions. In the case of an NVOC-protected amino acid, the degradation product is a nitrosobenzaldehyde, while the degradation product for

the other is a nitrosophenyl ketone. For instance, it is believed that the product aldehyde from NVOC degradation reacts with free amines to form a Schiff base (imine) that affects the remaining polymer synthesis. Preferred photoremovable protecting groups react slowly or reversibly with the oligomer on the support.

Again not wishing to be bound by theory, it is believed that the product ketone from irradiation of a MeNVOC-protected oligomer reacts at a slower rate with nucleophiles on the oligomer than the product aldehyde from irradiation of the same NVOC-protected oligomer. Although not unambiguously determined, it is believed that this difference in reaction rate is due to the difference in general reactivity between aldehyde and ketones towards nucleophiles due to steric and electronic effects.

The photoremovable protecting groups of the present invention are readily removed. For example, the photolysis of N-protected L-phenylalanine in solution and having different photoremovable protecting groups was analyzed, and the results are presented in the following table:

Table

Photolysis of Protected L-Phe-OH

Solvent	$t_{1/2}$ in seconds			
	NBOC	NVOC	MeNVOC	MeNPOC
Dioxane	1288	110	24	19
5mM H_2SO_4/Dioxane	1575	98	33	22

The half life, $t_{1/2}$, is the time in seconds required to remove 50% of the starting amount of protecting group. NBOC is the 6-nitrobenzyloxycarbonyl group, NVOC is the 6-nitroveratryloxycarbonyl group, MeNVOC is the methyl-6-nitroveratryloxycarbonyl group,

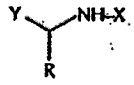
and MeNPOC is the methyl-6-nitropiperonyloxycarbonyl group. The photolysis was carried out in the indicated solvent with 362/364 nm-wavelength irradiation having an intensity of 10 mW/cm^2, and the concentration of each protected phenylalanine was 0.10 mM.

The table shows that deprotection of NVOC-, MeNVOC-, and MeNPOC-protected phenylalanine proceeded faster than the deprotection of NBOC. Furthermore, it shows that the deprotection of the two derivatives that are substituted on the benzylic carbon, MeNVOC and MeNPOC, were photolyzed at the highest rates in both dioxane and acidified dioxane.

1. Use of Photoremovable Groups During Solid-Phase Synthesis of Peptides

The formation of peptides on a solid-phase support requires the stepwise attachment of an amino acid to a substrate-bound growing chain. In order to prevent unwanted polymerization of the monomeric amino acid under the reaction conditions, protection of the amino terminus of the amino acid is required. After the monomer is coupled to the end of the peptide, the N-terminal protecting group is removed, and another amino acid is coupled to the chain. This cycle of coupling and deprotecting is continued for each amino acid in the peptide sequence. See Merrifield, J. Am. Chem. Soc. (1963) 85:2149, and Atherton et al., "Solid Phase Peptide Synthesis" 1989, IRL Press, London, both incorporated herein by reference for all purposes. As described above, the use of a photoremovable protecting group allows removal of selected portions of the substrate surface, via patterned irradiation, during the deprotection cycle of the solid phase synthesis. This selectively allows spatial control of the synthesis--the next amino acid is coupled only to the irradiated areas.

In one embodiment, the photoremovable protecting groups of the present invention are attached to an activated ester of an amino acid at the amino terminus:



where R is the side chain of a natural or unnatural amino acid, X is a photoremovable protecting group, and Y is an activated carboxylic acid derivative. The photoremovable protecting group, X, is preferably NVOC, NPOC, PyROC, MeNVOC, MeNPOC, and the like as discussed above. The activated ester, Y, is preferably a reactive derivative having a high coupling efficiency, such as an acyl halide, mixed anhydride, N-hydroxysuccinimide ester, perfluorophenyl ester, or urethane protected acid, and the like. Other activated esters and reaction conditions are well known (See Atherton et al.).

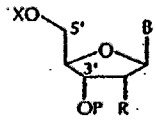
2. Use of Photoremovable Groups During Solid-Phase Synthesis of Oligonucleotides

The formation of oligonucleotides on a solid-phase support requires the stepwise attachment of a nucleotide to a substrate-bound growing oligomer. In order to prevent unwanted polymerization of the monomeric nucleotide under the reaction conditions, protection of the 5'-hydroxyl group of the nucleotide is required. After the monomer is coupled to the end of the oligomer, the 5'-hydroxyl protecting group is removed, and another nucleotide is coupled to the chain. This cycle of coupling and deprotecting is continued for each nucleotide in the oligomer sequence. See Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London, incorporated herein by reference for

all purposes. As described above, the use of a photoremovable protecting group allows removal, via patterned irradiation, of selected portions of the substrate surface during the deprotection cycle of the solid phase synthesis. This selectively allows spatial control of the synthesis--the next nucleotide is coupled only to the irradiated areas.

Oligonucleotide synthesis generally involves coupling an activated phosphorous derivative on the 3'-hydroxyl group of a nucleotide with the 5'-hydroxyl group of an oligomer bound to a solid support. Two major chemical methods exist to perform this coupling: the phosphate-triester and phosphoramidite methods (See Gait). Protecting groups of the present invention are suitable for use in either method.

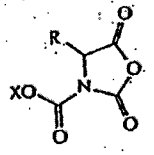
In a preferred embodiment, a photoremovable protecting group is attached to an activated nucleotide on the 5'-hydroxyl group:



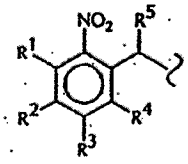
where B is the base attached to the sugar ring; R is a hydrogen atom when the sugar is deoxyribose or R is a hydroxyl group when the sugar is ribose; P represents an activated phosphorous group; and X is a photoremovable protecting group. The photoremovable protecting group, X, is preferably NV, NP, PyR, MeNV, MeNP, and the like as described above. The activated phosphorous group, P, is preferably a reactive derivative having a high coupling efficiency, such as a phosphate-triester, phosphoramidite or the like. Other activated phosphorous derivatives, as well as reaction conditions, are well known (See Gait).

E. Amino Acid N-Carboxy Anhydrides Protected With a Photoremovable Group

During Merrifield peptide synthesis, an activated ester of one amino acid is coupled with the free amino terminus of a substrate-bound oligomer. Activated esters of amino acids suitable for the solid phase synthesis include halo-formate, mixed anhydride, imidazoyl formate, acyl halide, and also includes formation of the activated ester in situ and the use of common reagents such as DCC and the like (See Atherton et al.). A preferred protected and activated amino acid has the general formula:



where R is the side chain of the amino acid and X is a photoremovable protecting group. This compound is a urethane-protected amino acid having a photoremovable protecting group attach to the amine. A more preferred activated amino acid is formed when the photoremovable protecting group has the general formula:

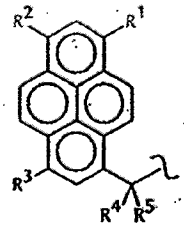


where R^1, R^2, R^3, and R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents (i.e., R^1-R^2, R^2-R^3, R^3-R^4) are substituted oxygen groups

that together form a cyclic acetal or ketal; and R^5 is a hydrogen atom, a alkoxyl, alkyl, hydrogen, halo, aryl, or alkenyl group.

A preferred activated amino acid is formed when the photoremovable protecting group is 6-nitroveratryloxycarbonyl. That is, R^1 and R^4 are each a hydrogen atom, R^2 and R^3 are each a methoxy group, and R^5 is a hydrogen atom. Another preferred activated amino acid is formed when the photoremovable group is 6-nitropiperonyl: R^1 and R^4 are each a hydrogen atom, R^2 and R^3 together form a methylene acetal, and R^5 is a hydrogen atom. Other protecting groups are possible. Another preferred activated ester is formed when the photoremovable group is methyl-6-nitroveratryl or methyl-6-nitropiperonyl.

Another preferred activated amino acid is formed when the photoremovable protecting group has the general formula:



where R^1, R^2, and R^3 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfanates, sulfido or phosphido group, and R^4 and R^5 independently are a hydrogen atom, an alkoxy, alkyl, halo, aryl, hydrogen, or alkenyl group. The resulting compound is a urethane-protected amino acid having a pyrenylmethyloxycarbonyl protecting group attached to the amine. A more preferred embodiment is formed when R^1 through R^5 are each a hydrogen atom.

The urethane-protected amino acids having a photoremovable protecting group of the present invention

are prepared by condensation of an N-protected amino acid with an acylating agent such as an acyl halide, anhydride, chloroformate and the like (See Fuller et al., U.S. Patent No. 4,946,942 and Fuller et al., J. Amer. Chem. Soc. (1990) 112:7414-7416, both herein incorporated by reference for all purposes).

Urethane-protected amino acids having photoremovable protecting groups are generally useful as reagents during solid-phase peptide synthesis, and because of the spatially selectivity possible with the photoremovable protecting group, are especially useful for the spatially addressable peptide synthesis. These amino acids are difunctional: the urethane group first serves to activate the carboxy terminus for reaction with the amine bound to the surface and, once the peptide bond is formed, the photoremovable protecting group protects the newly formed amino terminus from further reaction. These amino acids are also highly reactive to nucleophiles, such as deprotected amines on the surface of the solid support, and due to this high reactivity, the solid-phase peptide coupling times are significantly reduced, and yields are typically higher.

IV. Data Collection

A. Data Collection System

Substrates prepared in accordance with the above description are used in one embodiment to determine which of the plurality of sequences thereon bind to a receptor of interest. Fig. 11 illustrates one embodiment of a device used to detect regions of a substrate which contain fluorescent markers. This device would be used, for example, to detect the presence or absence of a labeled receptor such as an antibody which has bound to a synthesized polymer on a substrate.

Light is directed at the substrate from a light source 1002 such as a laser light source of the type well known to those of skill in the art such as a model no.

2025 made by Spectra Physics. Light from the source is directed at a lens 1004 which is preferably a cylindrical lens of the type well known to those of skill in the art. The resulting output from the lens 1004 is a linear beam rather than a spot of light, resulting in the capability to detect data substantially simultaneously along a linear array of pixels rather than on a pixel-by-pixel basis. It will be understood that which a cylindrical lens is used herein as an illustration of one technique for generating a linear beam of light on a surface, but it will be understood that other techniques could also be utilized.

The beam from the cylindrical lens is passed through a dichroic mirror or prism (1006) and directed at the surface of the suitably prepared substrate 1008. Substrate 1008 is placed on an x-y translation stage 1009 such as a model no. PM500-8 made by Newport. Light at certain locations on the substrate will be fluoresced and transmitted along the path indicated by dashed lines back through the dichroic mirror, and focused with a suitable lens 1010 such as an f/1.4 camera lens on a linear detector 1012 via a variable f stop focusing lens 1014. Through use of a linear light beam, it becomes possible to generate data over a line of pixels (such as about 1 cm) along the substrate, rather than from individual points on the substrate. In alternative embodiments, light is directed at a 2-dimensional area of the substrate and fluoresced light detected by a 2-dimensional CCD array. Linear detection is preferred because substantially higher power densities are obtained.

Detector 1012 detects the amount of light fluoresced from the substrate as a function of position. According to one embodiment the detector is a linear CCD array of the type commonly known to those of skill in the art. The x-y translation stage, the light source, and the detector 1012 are all operably connected to a

computer 1016 such as a IBM PC-AT or equivalent for control of the device and data collection from the CCD array.

In operation, the substrate is appropriately positioned by the translation stage. The light source is then illuminated, and intensity data are gathered with the computer via the detector.

Fig. 12 illustrates the architecture of the data collection system in greater detail. Operation of the system occurs under the direction of the photon counting program 1102 (photon), included herewith as Appendix B. The user inputs the scan dimensions, the number of pixels or data points in a region, and the scan speed to the counting program. Via a GPIB bus 1104 the program (in an IBM PC compatible computer, for example) interfaces with a multichannel scaler 1106 such as a Stanford Research SR 430 and an x-y stage controller 1108 such as a PM500. The signal from the light from the fluorescing substrate enters a photon counter 1110, providing output to the scaler 1106. Data are output from the scaler indicative of the number of counts in a given region. After scanning a selected area, the stage controller is activated with commands for acceleration and velocity, which in turn drives the scan stage 1112 such as a PM500-A to another region.

Data are collected in an image data file 1114 and processed in a scaling program 1116, also included in Appendix B. A scaled image is output for display on, for example, a VGA display 1118. The image is scaled based on an input of the percentage of pixels to clip and the minimum and maximum pixel levels to be viewed. The system outputs for use the min and max pixel levels in the raw data.

B. Data Analysis

The output from the data collection system is an array of data indicative of fluorescent intensity versus location on the substrate. The data are typically taken over regions substantially smaller than the area in which synthesis of a given polymer has taken place. Merely by way of example, if polymers were synthesized in squares on the substrate having dimensions of 500 microns by 500 microns, the data may be taken over regions having dimensions of 5 microns by 5 microns. In most preferred embodiments, the regions over which ~~florescence~~ fluorescence data are taken across the substrate are less than about 1/2 the area of the regions in which individual polymers are synthesized, preferably less than 1/10 the area in which a single polymer is synthesized, and most preferably less than 1/100 the area in which a single polymer is synthesized. Hence, within any area in which a given polymer has been synthesized, a large number of fluorescence data points are collected.

A plot of number of pixels versus intensity for a scan of a cell when it has been exposed to, for example, a labeled antibody will typically take the form of a bell curve, but spurious data are observed, particularly at higher intensities. Since it is desirable to use an average of fluorescent intensity over a given synthesis region in determining relative binding affinity, these spurious data will tend to undesirably skew the data.

Accordingly, in one embodiment of the invention the data are corrected for removal of these spurious data points, and an average of the data points is thereafter utilized in determining relative binding efficiency.

Fig. 13 illustrates one embodiment of a system for removal of spurious data from a set of fluorescence data such as data used in affinity screening studies. A user or the system inputs data relating to the chip location and cell corners at step 1302. From this

information and the image file, the system creates a computer representation of a histogram at step 1304, the histogram (at least in the form of a computer file) plotting number of data pixels versus intensity.

For each cell, a main data analysis loop is then performed. For each cell, at step 1306, the system calculates the total intensity or number of pixels for the bandwidth centered around varying intensity levels. For example, as shown in the plot to the right of step 1306, the system calculates the number of pixels within the band of width w. The system then "moves" this bandwidth to a higher center intensity, and again calculates the number of pixels in the bandwidth. This process is repeated until the entire range of intensities has been scanned, and at step 1308 the system determines which band has the highest total number of pixels. The data within this bandwidth are used for further analysis. Assuming the bandwidth is selected to be reasonably small, this procedure will have the effect of eliminating spurious data located at the higher intensity levels. The system then repeats at step 1310 if all cells have been evaluated, or repeats for the next cell.

At step 1312 the system then integrates the data within the bandwidth for each of the selected cells, sorts the data at step 1314 using the synthesis procedure file, and displays the data to a user on, for example, a video display or a printer.

V. Representative Applications

A. Oligonucleotide Synthesis

The generality of light directed spatially addressable parallel chemical synthesis is demonstrated by application to nucleic acid synthesis.

Example

Light activated formation of a thymidine-cytidine dimer was carried out. A three dimensional representation of a fluorescence scan showing a checkerboard pattern generated by the light-directed synthesis of a dinucleotide is shown in Fig. 8. 5'-nitroveratryl thymidine was attached to a synthesis substrate through the 3' hydroxyl group. The nitroveratryl protecting groups were removed by illumination through a 500 mm checkerboard mask. The substrate was then treated with phosphoramidite activated 2'-deoxycytidine. In order to follow the reaction fluorometrically, the deoxycytidine had been modified with an FMOC protected aminohexyl linker attached to the exocyclic amine (5'-O-dimethoxytrityl-4-N-(6-N-fluorenylmethylcarbamoyl-hexylcarboxy)-2'-deoxycytidine). After removal of the FMOC protecting group with base, the regions which contained the dinucleotide were fluorescently labelled by treatment of the substrate with 1 mM FITC in DMF for one hour.

The three-dimensional representation of the fluorescent intensity data in Fig. 14 clearly reproduces the checkerboard illumination pattern used during photolysis of the substrate. This result demonstrates that oligonucleotides as well as peptides can be synthesized by the light-directed method.

VI. Conclusion

The inventions herein provide a new approach for the simultaneous synthesis of a large number of compounds. The method can be applied whenever one has chemical building blocks that can be coupled in a solid-phase format, and when light can be used to generate a reactive group.

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review

of this disclosure. Merely by way of example, while the invention is illustrated primarily with regard to peptide and nucleotide synthesis, the invention is not so limited. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

82
60

WHAT IS CLAIMED IS:

1. A reactor system for synthesizing a plurality of polymer sequences on a substrate comprising:

a) a reactor for contacting reaction fluids to said substrate;

b) a system for delivering selected reaction fluids to said reactor;

c) a translation stage for moving a mask or substrate from at least a first relative location relative to a second relative location;

d) a light for illuminating said substrate through a mask at selected times; and

e) an appropriately programmed digital computer for selectively directing a flow of fluids from said reactor system, selectively activating said translation stage, and selectively illuminating said substrate so as to form a plurality of diverse polymer sequences on said substrate at predetermined locations.

2. The reactor system as recited in claim 1 adapted to provide a plurality of monomers in a reaction fluid to said substrate, said substrate used for an initial screening of polymer sequences.

3. An ordered method for forming a plurality of polymer sequences by sequential addition of reagents comprising the step of serially protecting and deprotecting portions of said plurality of polymer sequences for addition of other portions of said polymer sequences using a binary synthesis strategy.

4. The method as recited in claim 3 wherein said binary synthesis strategy is a binary masking strategy.

5. The method as recited in claim 4 wherein said masking strategy in which said masking strategy provides

83
62

at least two consecutive steps in which a mask factors a previous mask by protecting a portion of a previously illuminated portions to light and exposing a portion of a previously protected portions to light.

6. The method as recited in claim 4 in which said masking strategy in which at least two successive steps in said masking strategy illuminate about one half of a region of interest on said substrate.

7. The method as recited in claim 4 wherein said masking strategy forms a plurality of polymer sequences on a single substrate.

8. The method as recited in claim 4 wherein said masks are arranged in a gray code masking scheme, said gray code masking scheme having one edge illumination on each of a plurality of synthesis sites.

9. The method as recited in claim 4 wherein said masking scheme results in a minimum number of masking steps for a number of polymers synthesized.

10. The method as recited in claim 4 wherein all possible polymers of length l are formed with a given basis set of monomers.

11. The method as recited in claim 4 wherein said masking strategy is developed in an appropriately programmed digital computer inputting at least a desired basis set, and length of polymers.

12. The method as recited in claim 4 wherein all possible polymers of a length less than or equal to l are formed with a given basis set of monomers.

13. The method as recited in claim 4 further comprising the step of forming a portion of said polymers with a non-binary masking scheme.

14. The method as recited in claim 10 further comprising the step of outputting a masking strategy.

15. The method as recited in claim 10 further comprising the step of outputting a map of synthesized polymers on said substrate.

16. The method as recited in claim 15 wherein said map is in the form of Fig. 10.

17. A method of screening a plurality of linker polymers for use in binding affinity studies comprising the steps of:

a) forming a plurality of linker polymers on a substrate in selected regions, said linker polymers formed by the steps of recursively:

i) on a surface of a substrate, irradiating a portion of said selected regions to remove a protective group; and

ii) contacting said surface with a monomer;

b) contacting said plurality of linker polymers with a ligand; and

c) contacting said ligand with a labeled receptor.

18. The method as recited in claim 17 wherein said ligand is a polypeptide.

19. The method as recited in claim 17 wherein said receptor is an antibody.

20. The method as recited in claim 17 wherein said monomers added in step ii) are the same in each of said

recursive steps, said selected regions comprising linker molecules of different lengths.

21. The method as recited in claim 17 wherein said labelled receptor is a fluoresceinated receptor.

22. A system for determining affinity of a receptor to a ligand comprising:

a) means for applying light to a surface of a substrate, said substrate comprising a plurality of ligands at predetermined locations, said means for applying directing light providing simultaneous illumination at a plurality of said predetermined locations; and

b) an array of detectors for detecting light fluoresced at said plurality of predetermined locations.

23. A system as recited in claim 22 wherein said means for applying light comprises a point light source and a cylindrical lens for focusing said point light source along a substantially linear path.

24. A system as recited in claim 22 wherein said array of detectors comprises a linear array.

25. A system as recited in claim 22 wherein said array of detectors comprises a linear CCD array.

26. In a digital computer, a method of determining the tendency of a receptor to bind to a ligand comprising:

a) exposing fluorescently labelled receptors to a substrate, said substrate comprising a plurality of ligands in regions at known locations;

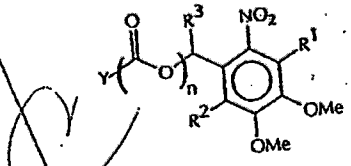
b) at a plurality of data collection points within each of said regions, determining an amount of light fluoresced from said data collection points;

c) removing said data collection points deviating from a preset amount from a predetermined statistical distribution; and

d) determining a relative binding affinity of said receptor to remaining data collection points.

27. The method as recited in claim 26 wherein said predetermined statistical distribution is a normal distribution.

28. A compound having the formula:



wherein n = 0 or 1; Y is selected from the group consisting of an oxygen of the carboxyl group of a natural or unnatural amino acid, an amino group of a natural or unnatural amino acid, or the C-5' oxygen group of a natural or unnatural deoxyribonucleic or ribonucleic acid; R^1 and R^2 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfido, or phosphido group; and R^3 is a alkoxy, alkyl, aryl, hydrogen, or alkenyl group.

29. The compound of claim 28 wherein Y is the C-5' oxygen group of a natural or unnatural deoxyribonucleic or ribonucleic acid.

30. The compound of claim 29 wherein n = 0.

31. The compound of claim 29 wherein R^1 and R^2 are each a hydrogen atom.

87
88

32. The compound of claim 31 wherein R^3 is a hydrogen atom.

33. The compound of claim 31 wherein R^3 is a methyl group.

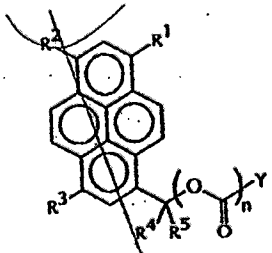
34. The compound of claim 28 wherein Y is an oxygen of the carboxyl group of an amino acid and n = 0.

35. The compound of claim 34 wherein R^1 and R^2 are each a hydrogen atom.

36. The compound of claim 35 wherein R^3 is a hydrogen atom.

37. The compound of claim 35 wherein R^3 is a methyl group.

38. A compound having the formula:



wherein n = 0 or 1; Y is selected from the group consisting of an amino group of a natural or unnatural amino acid or the C-5' oxygen group of a natural or unnatural deoxyribonucleic and ribonucleic acid; R^1, R^2, and R^3 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfido or phosphido group; R^4 and R^5 independently are a

88.

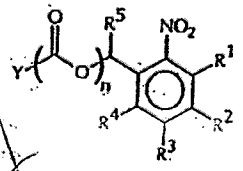
alkoxy, alkyl, hydrogen, halo, aryl, hydrogen, or alkenyl group.

39. The compound of claim 38 wherein R^1 through R^3 are each a hydrogen atom.

40. The compound of claim 39 wherein R^4 and R^5 are each a hydrogen atom.

41. The compound of claim 39 wherein R^4 and R^5 are each a methyl group.

42. A compound having the formula:



wherein n = 0 or 1; Y is a C-5' oxygen group of a natural or unnatural deoxyribonucleic and ribonucleic acid; R^1 through R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfido, or phosphido group; and R^5 is a alkoxy, alkyl, aryl, or alkenyl group.

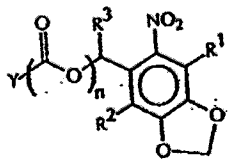
43. The compound of claim 42 wherein R^2 and R^3 are each a methoxy group.

44. The compound of claim 43 wherein R^1 and R^4 are each a hydrogen atom.

45. The compound of claim 44 wherein R^5 is a methyl group.

89
87

46. A compound having the formula:



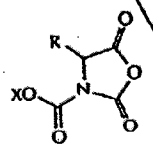
wherein n = 0 or 1; Y is an atom to be protected; R^1 and R^2 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfido, or phosphido group; and R^3 is a alkoxy, alkyl, aryl, or alkenyl group.

47. The compound of claim 46 wherein Y is selected from the group consisting of an oxygen of the carboxyl group of a natural or unnatural amino acid, or the C-5' oxygen group of a natural or unnatural deoxyribonucleic or ribonucleic acid, or the amino group of a natural or unnatural amino acid.

48. The compound of claim 47 wherein R^1 and R^2 are hydrogen.

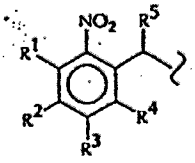
49. The compound of claim 48 wherein R^3 is a methyl group.

50. A compound having the formula:



where R is a side chain of a natural or unnatural amino acid and X is a photoremovable protecting group.

51. The compound of claim 50 wherein X has the following formula:



where R^1, R^2, R^3, and R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents are substituted oxygen groups that together form a cyclic acetal or ketal; and R^5 is a hydrogen atom, a alkoxyl, alkyl, hydrogen, halo, aryl, or alkenyl group.

52. The compound of claim 51 wherein R^1 and R^4 are each a hydrogen atom, and R^2 and R^3 are each a methoxy group.

53. The compound of claim 52 wherein R^5 is a methyl group.

54. The compound of claim 51 wherein R^2 and R^3 are substituted oxygen groups that together form a cyclic acetal.

55. The compound of claim 54 wherein R^1 and R^4 are each a hydrogen atom.

56. The compound of claim 55 wherein R^5 is a methyl group.

add C4

add D2

Print Of Drawings
As Originally Filed

07 624120
14-1

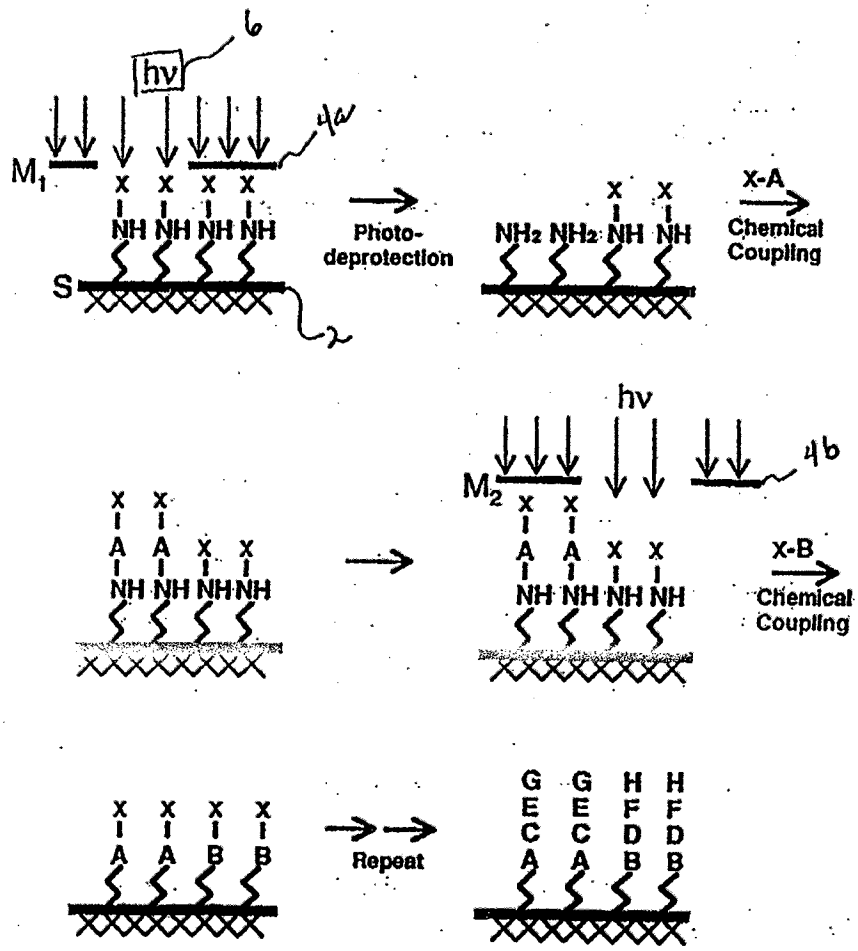


Figure 1

Print Of Drawings
As Originally Filed

07 624120

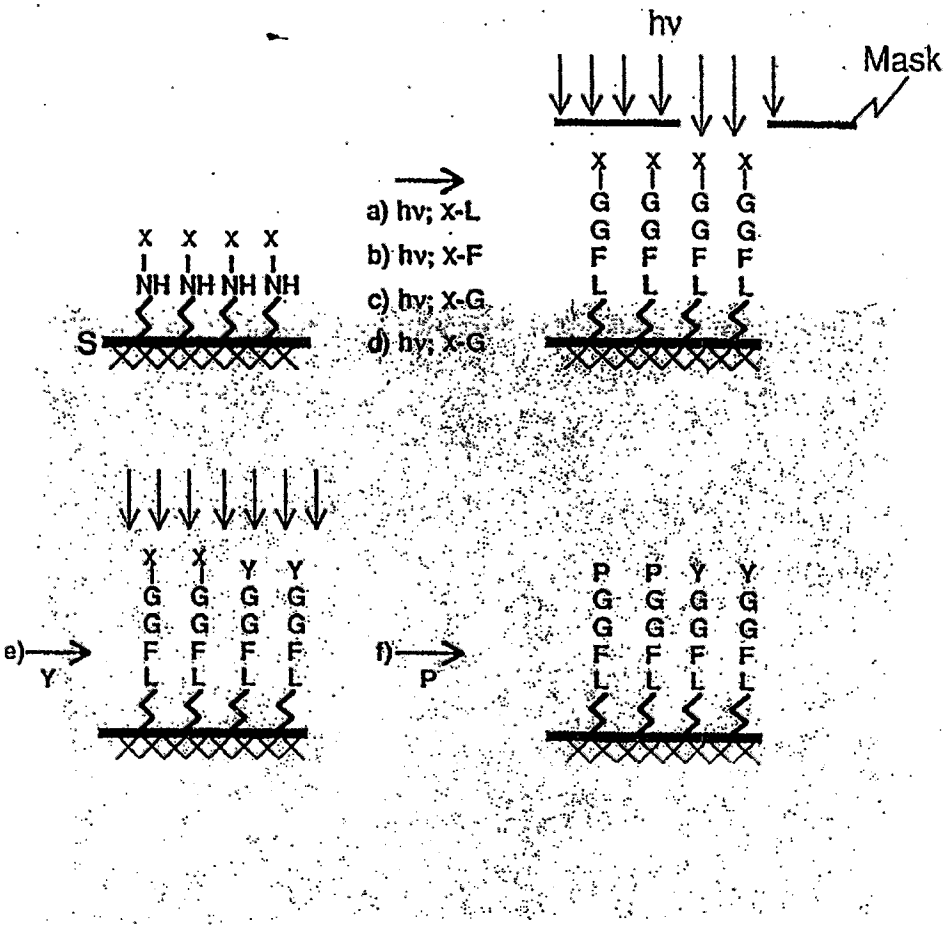


Figure 2

Print Of Drawings
As Originally Filed

07 624120

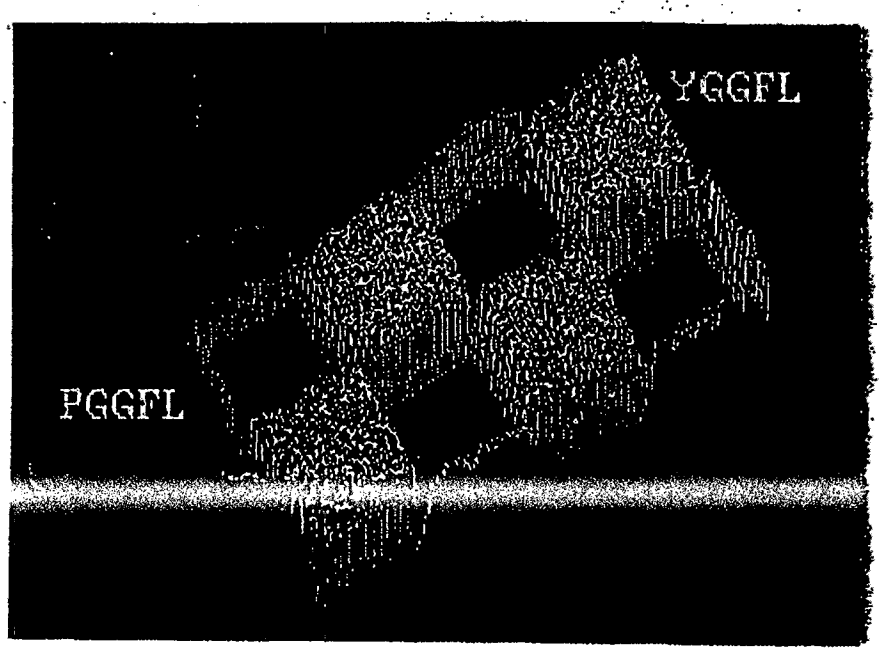


Figure 3

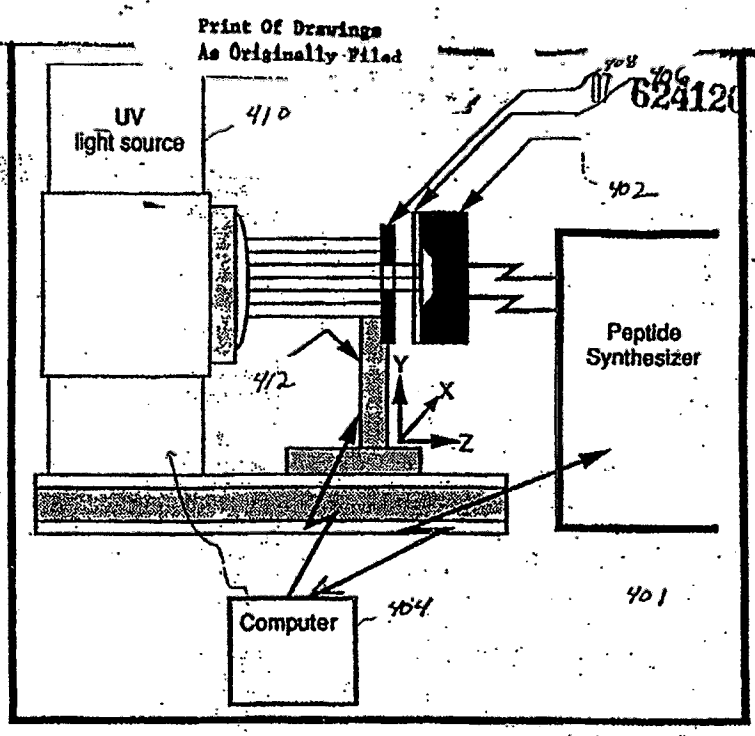

Print Of Drawings
As Originally Filed
UV
light source
410
408
406
624120
402
Peptide
Synthesizer
412
Y
X
Z
Computer
404
401

FIG. 4

Print Of Drawings
As Originally Filed

07 624120

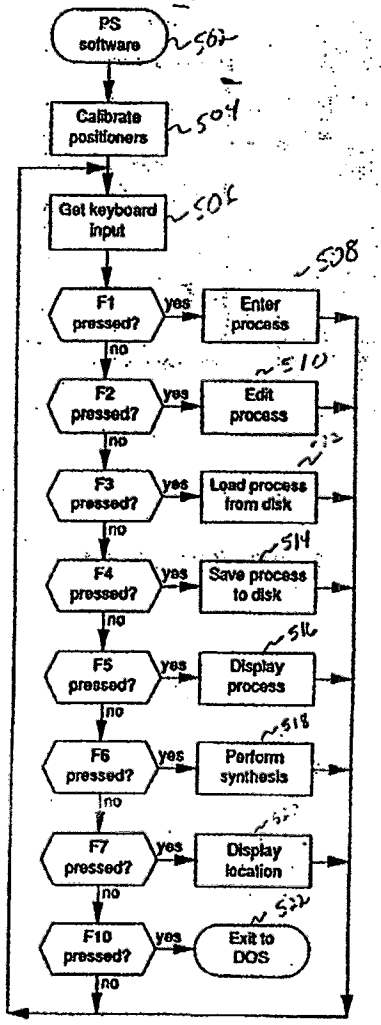


Fig. 5a

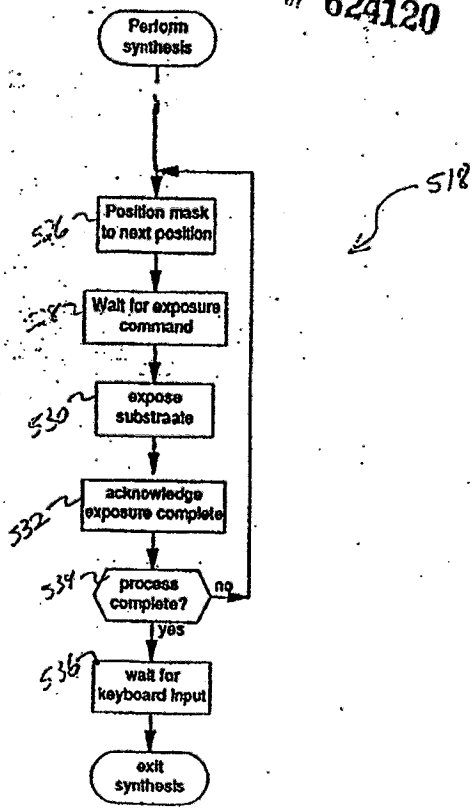


Fig. 5b

-1-

Print Of Drawings
As Originally Filed

07 624120

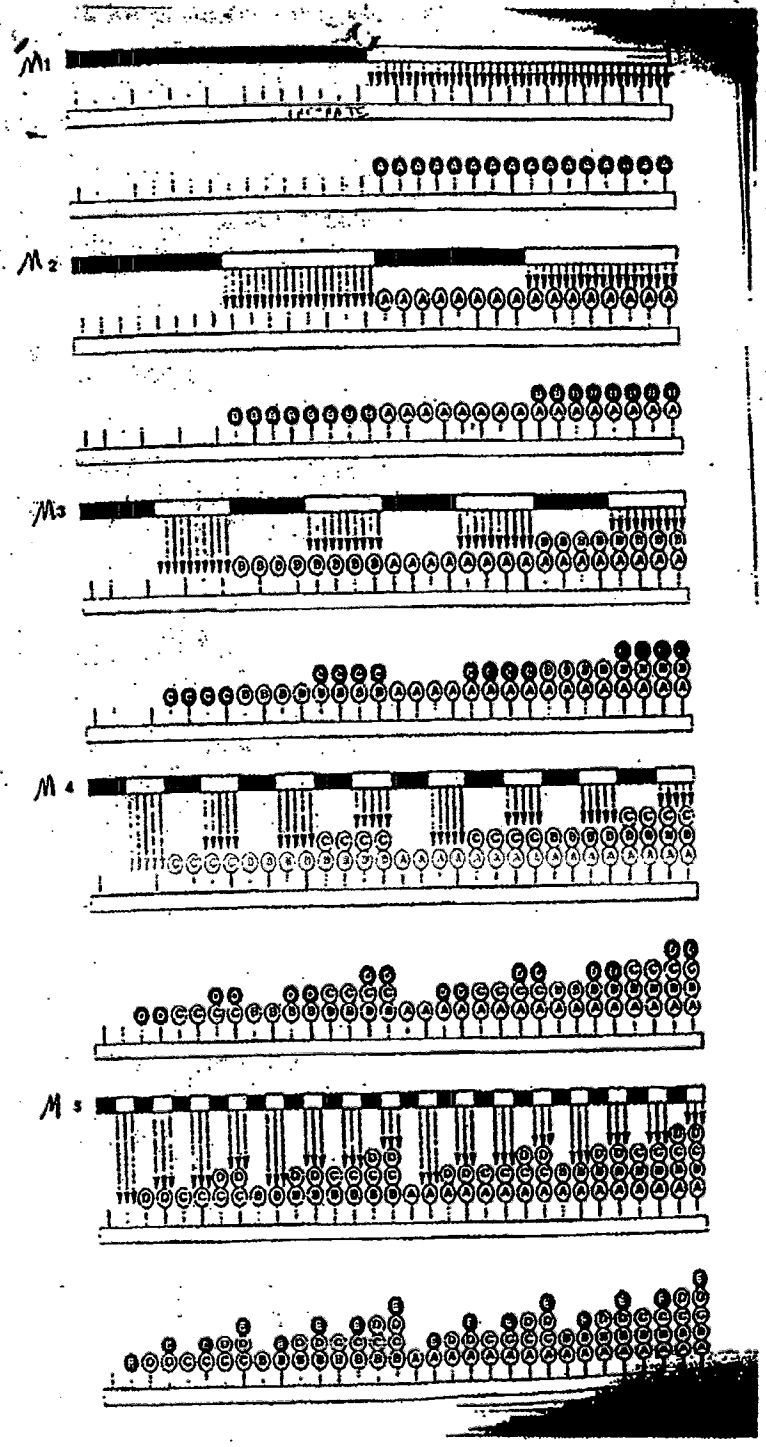


FIG. 6

Print Of Drawings
As Originally Filed

07 624120

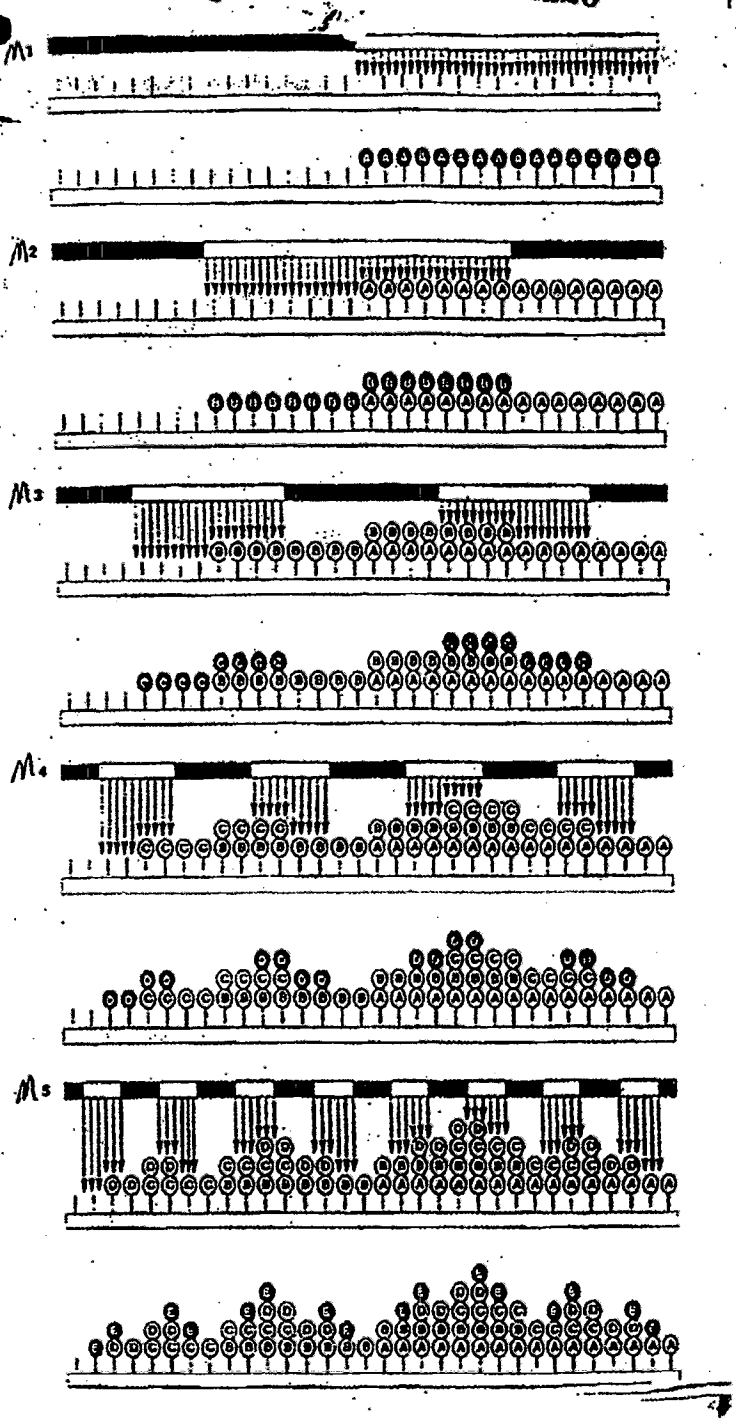


FIG. 7

Print Of Drawings
As Originally Filed

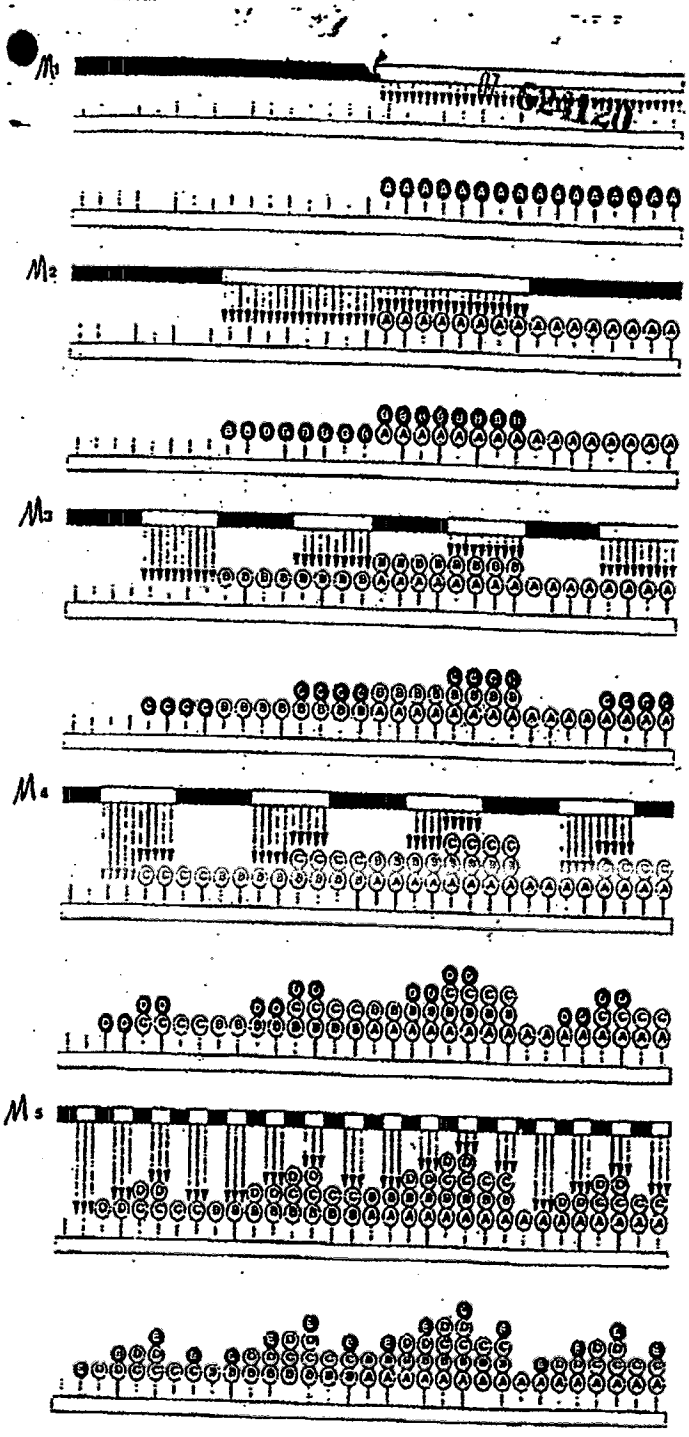


Fig. 8

Print Of Drawings
As Originally Filed

07 624120

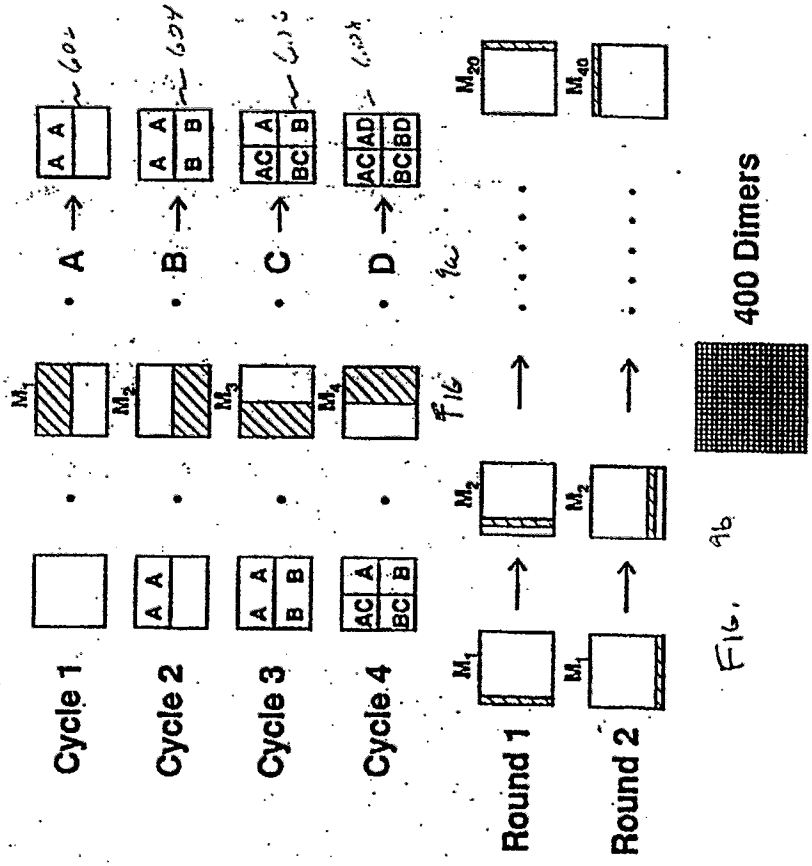


Fig. 9a

Fig. 9b

Print Of Drawings
As Originally Filed

07 624120

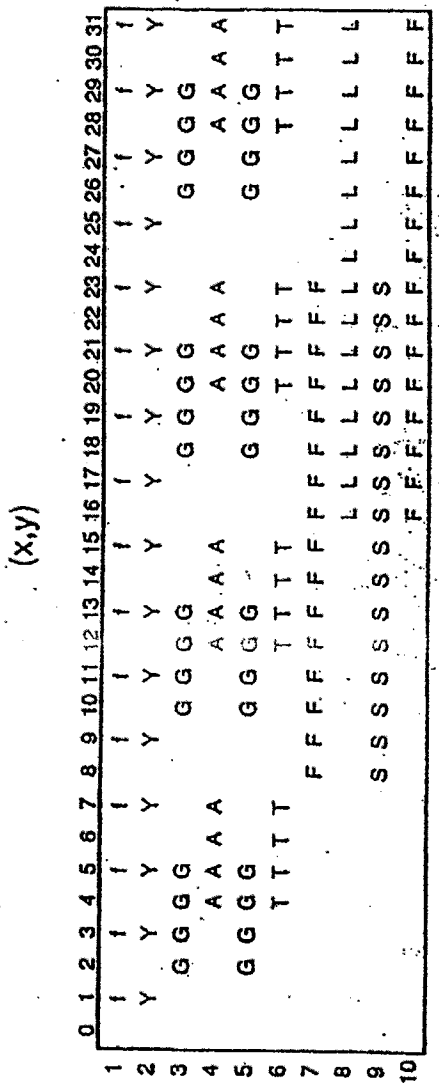
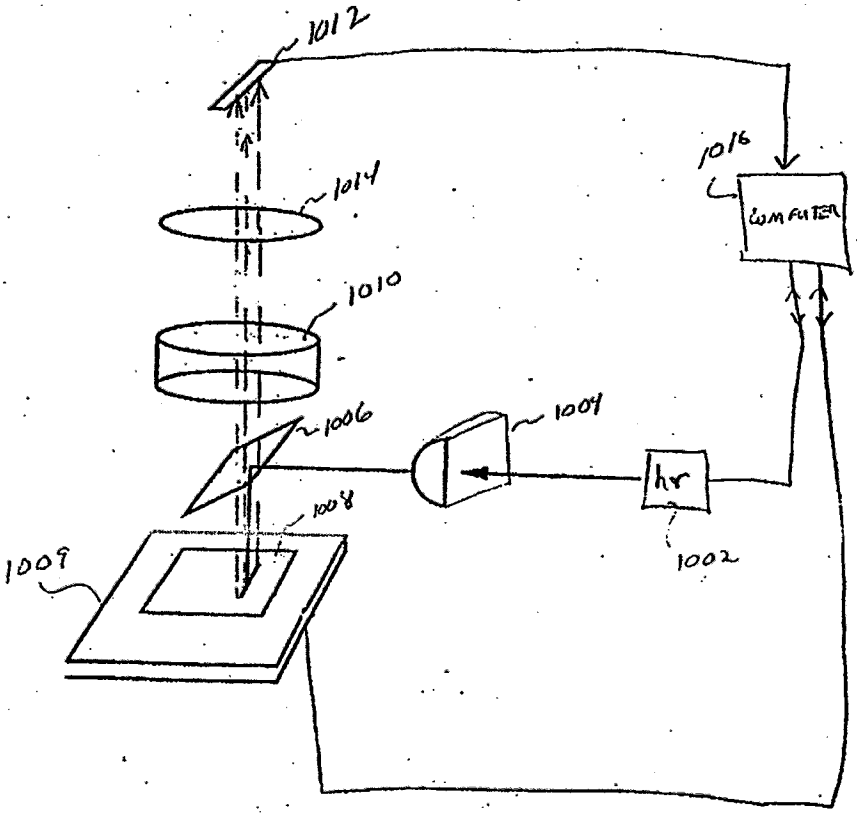


FIG. 15

Print Of Drawings
As Originally Filed

07 624120



Figure/1

Print Of Drawings
As Originally Filed

07 624120

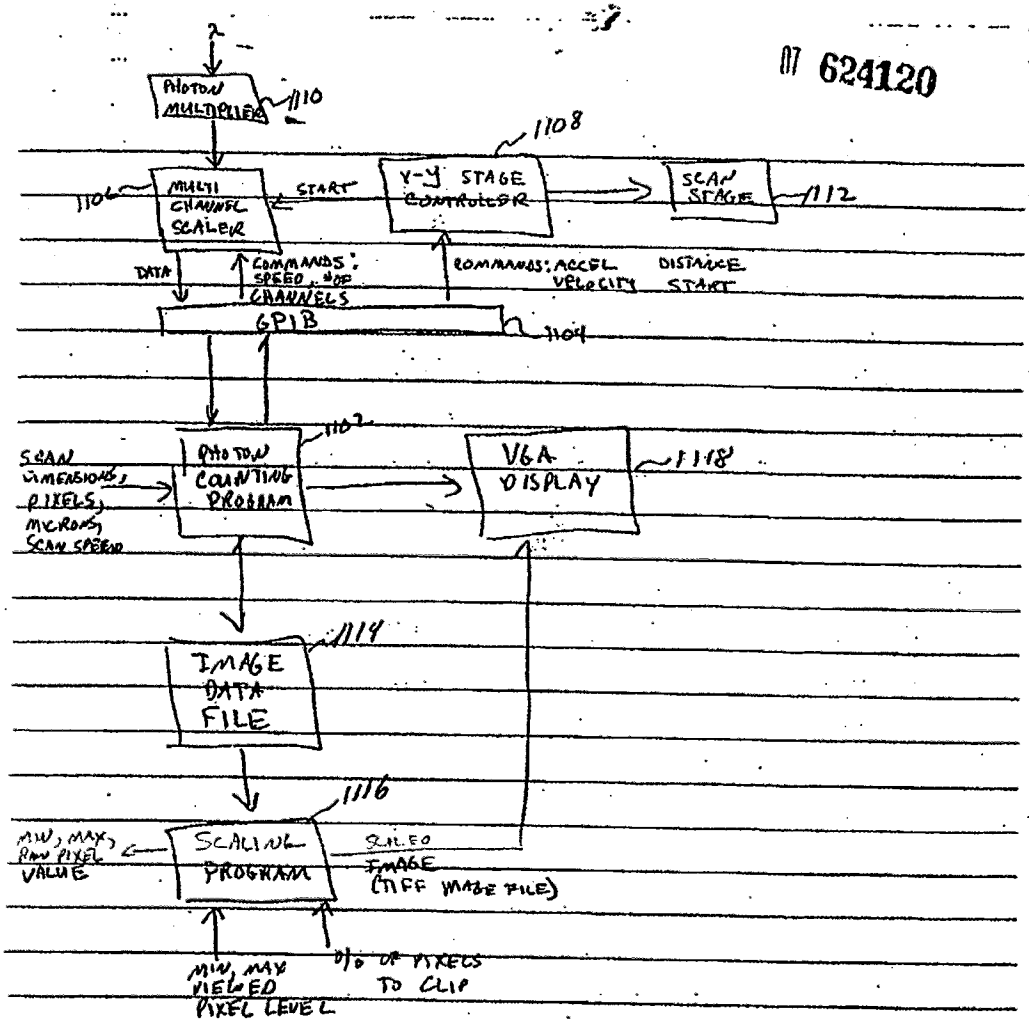


FIG. 12

Print Of Drawings
As Originally Filed

624120

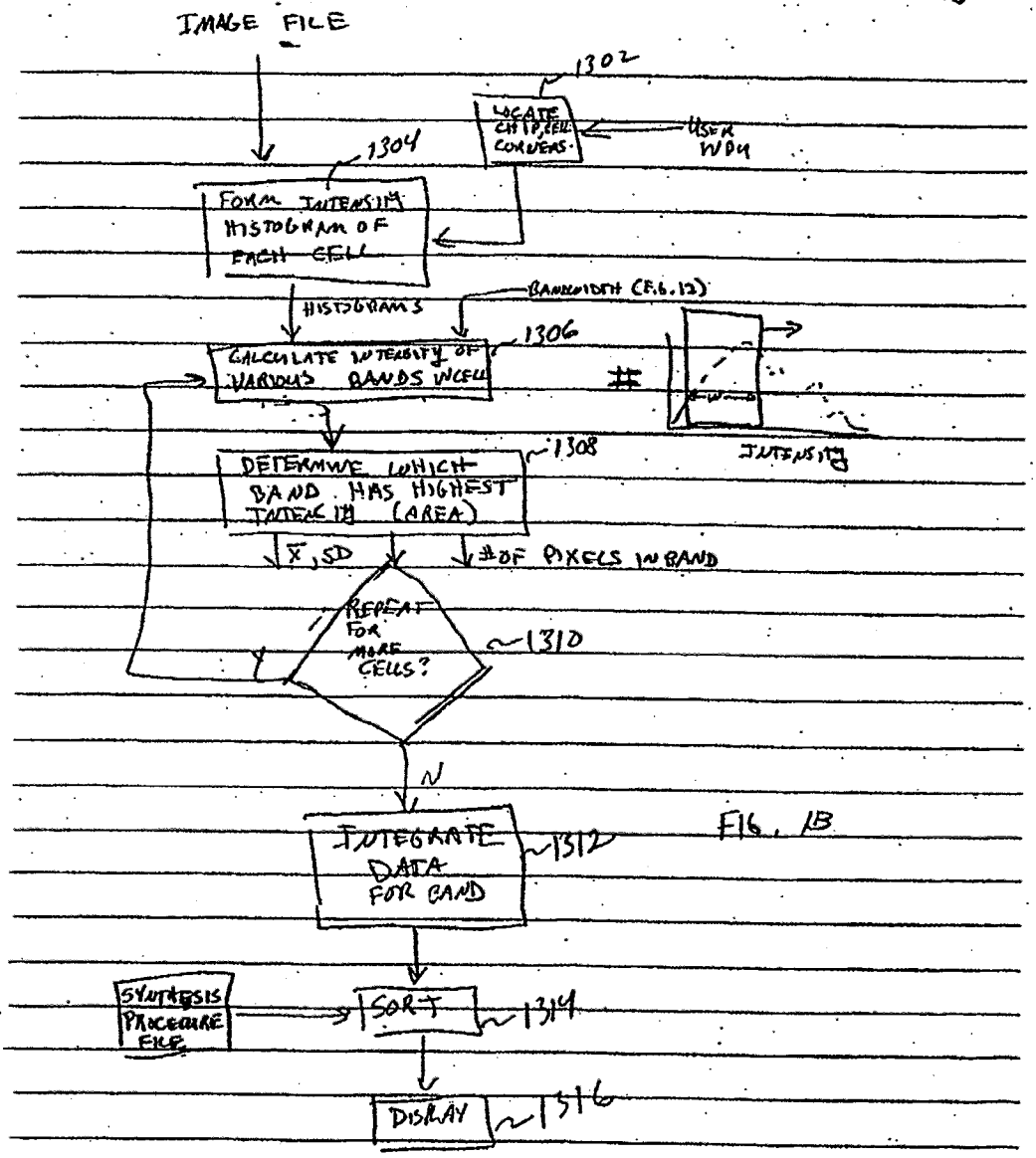


FIG. 13

Print Of Drawings
As Originally Filed

07 624120

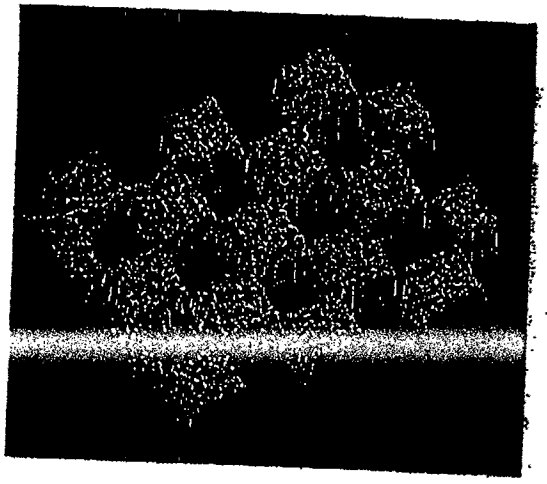


FIG. 14

PATENT APPLICATION SERIAL NO. 08 082937

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

PTO-1556
(5/87)

11509-101

PATENT APPLICATION

HYBRIDIZATION AND SEQUENCING OF NUCLEIC ACIDS

Inventor(s):

STEPHEN P.A. FODOR, a citizen of the United States of America, residing at 3863 Nathan Way, Palo Alto, CA 94303;

ROBERT J. LIPSHUTZ, a citizen of the United States of America, residing at 970 Palo Alto Avenue, Palo Alto, CA 94301; and

XIAOHUA HUANG, a citizen of the People's Republic of China, residing at 937 Jackson Street, Mt. View, CA 94043.

Assignee:

AFFYMAX TECHNOLOGIES N.V.
De Ruyderkade 62
Curacao, Netherlands Antilles

A Netherlands Antilles corporation.

Entity Status:

Large

TOWNSEND and TOWNSEND KHOURIE and CREW
Steuart Street Tower
One Market Plaza
San Francisco, CA 94105
(415) 326-2400

08/505,919

11509-101

HYBRIDIZATION AND SEQUENCING OF NUCLEIC ACIDS

ABSTRACT OF THE DISCLOSURE

Devices and techniques for hybridization of nucleic acids and for determining the sequence of nucleic acids. Arrays of nucleic acids are formed by techniques, preferably high resolution, light-directed techniques. Positions of hybridization of a target nucleic acid are determined by, e.g., epifluorescence microscopy. Devices and techniques are proposed to determine the sequence of a target nucleic acid more efficiently and more quickly through such synthesis and detection techniques.

11509-101

HYBRIDIZATION AND SEQUENCING OF NUCLEIC ACIDS

GOVERNMENT RIGHTS

The invention described herein arose in the course of or under Contract No. DE-FG03-92ER81275 (Grant No. 21012-92-II) between the Department of Energy and Affymax; and in the course of or under NIH Contract No. 1R01HG00813-01.

BACKGROUND OF THE INVENTION

The present invention relates to the field of nucleic acid analysis, detection, and sequencing. More specifically, in one embodiment the invention provides improved techniques for synthesizing arrays of nucleic acids, hybridizing nucleic acids, detecting mismatches in a double-stranded nucleic acid composed of a single-stranded probe and a target nucleic acid, and determining the sequence of DNA or RNA or other polymers.

It is important in many fields to determine the sequence of nucleic acids because, for example, nucleic acids encode the enzymes, structural proteins, and other effectors of biological functions. In addition to segments of nucleic acids that encode polypeptides, there are many nucleic acid sequences involved in control and regulation of gene expression.

The human genome project is one example of a project using nucleic acid sequencing techniques. This project is directed toward determining the complete sequence of the genome of the human organism. Although such a sequence would not necessarily correspond to the sequence of any specific individual, it will provide significant information as to the general organization and specific sequences contained within genomic segments from particular individuals. The human genome project will also provide mapping information useful for further detailed studies.

The need for highly rapid, accurate, and inexpensive sequencing technology is nowhere more apparent than in a demanding sequencing project such as the human genome project. To complete the sequencing of a human genome will require the determination of approximately 3×10^9, or 3 billion, base pairs.

The procedures typically used today for sequencing include the methods described in Sanger et al., Proc. Natl. Acad. Sci. USA (1977) 74:5463-5467, and Maxam et al., Methods in Enzymology (1980) 65:499-559. The Sanger method utilizes enzymatic elongation with chain terminating dideoxy nucleotides. The Maxam and Gilbert method uses chemical reactions exhibiting base-specific cleavage reactions. Both methods require a large number of complex manipulations, such as isolation of homogeneous DNA fragments,